

# 온라인 커뮤니티 특화 감성 사전 구축을 위한 새로운 용어 극성값 분석 시스템

빅데이터미네이터

# 진행 사항

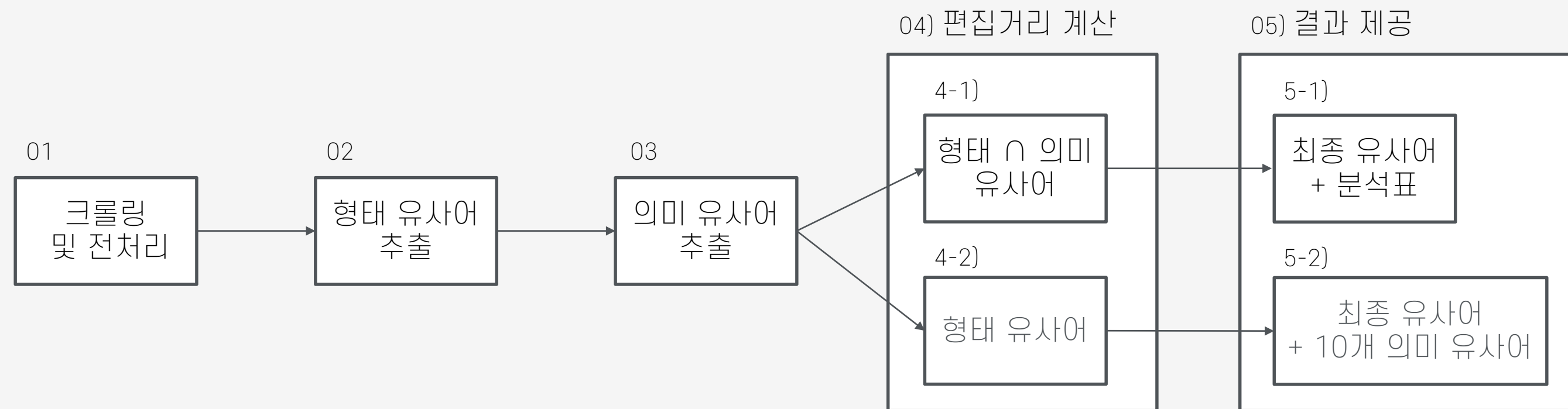
1. 학회 발표 동영상 및 드림패스 동영상 제출
2. 특허 아이디어 구체화 및 TEST

# 진행 사항 - 특허 아이디어 빌딩



## 💡 KEY POINT

⇒ 온라인 형태 변형어의 **편집거리 기반 형태 및 의미 유사어** 분석 시스템



- 1) “네이버 블로그” 에서 사용자가 입력한 형태 변형어를 검색하여 데이터 크롤링 및 전처리
- 2) 입력한 형태 변형어의 형태 유사어 추출 (문장 단위 & 자소 단위 학습)
- 3) 입력한 형태 변형어의 의미 유사어 추출
- 4) 추출된 유사어들의 편집거리 계산
- 5) 최종 유사어 및 편집거리 분석표 제공

- ## (예시) 형태 변형어 ‘댕댕이’ 검색



## 자소 단위 학습



# 진행 사항 - 특허 아이디어 빌딩

## 2) 입력한 형태 변형어의 형태 유사어 추출

### 2-1) 문장 단위 학습

#문장 학습 - 형태 유사어 추출

```
model1 = fasttext.load_model('data_result.bin')
model1.get_nearest_neighbors("댕댕이")
```

Warning : `load\_model` does not return WordVector

```
[(0.9838256239891052, '사실'),
 (0.9823348522186279, '강아지를'),
 (0.9820657968521118, '댕댕이는'),
 (0.9797086715698242, '댕댕이들'),
 (0.9775006175041199, '경우'),
 (0.9774555563926697, '강아지와'),
 (0.9769712686538696, '댕댕이가'),
 (0.9754970669746399, '강아지는'),
 (0.9742645025253296, '강아지'),
 (0.9709970355033875, '있는')]
```

### 2-2) 자소 단위 학습

```
1 from soynlp.hangle import compose, decompose, character_is_korean
2
3 #자소단위 학습 - 형태 유사어 추출
4 model3 = fasttext.load_model('re_result.bin')
5 model3.get_nearest_neighbors("댕댕이")
```

Warning : `load\_model` does not return WordVectorModel or SupervisedModel any more, milar.

```
[(0.32594582438468933, '크아오펬오펬오펬'),
 (0.25433990359306335, '나오펬오펬오펬'),
 (0.2458125650882721, '오펬오펬'),
 (0.1965695321559906, '오펬오펬'),
 (0.19008754193782806, '오펬오펬'),
 (0.16549181938171387, '자오펬오펬오펬'),
 (0.1537555456161499, '바오펬오펬'),
 (0.13369370996952057, '호오펬오펬오펬'),
 (0.12780514359474182, '기오펬오펬오펬'),
 (0.10005773603916168, '바오펬오펬오펬오펬오펬')]
```

```
[hgtk.text.compose(found) for found in [word for score, word in model3.get_nearest.
 ['카페에', '놀러', '명', '털', '및', '조금', '바로', '혼자', '계속', '반려견']]
```

## 3) 입력한 형태 변형어의 의미 유사어 추출

```
print(loader_model.wv.most_similar("댕댕이", topn=10))
```

```
[('댕댕이삼계탕', 0.9992862343788147), ('강아지삼계탕', 0.9992821216583252), ('강아지돛', 0.9992791414260864), ('퓨어
메라', 0.9992114901542664), ('이벤트', 0.9991117119789124), ('강아지간식', 0.999061107635498), ('강아지', 0.9990420937
538147), ('강아지돛만들기', 0.9990110397338867), ('중입니다.분', 0.9989457130432129), ('애견용품', 0.998928427696228)]
```

# 진행 사항 - 특허 아이디어 빌딩

## 4) 추출된 유사어들의 편집거리 계산

### 4-1) 의미 유사어가 형태 유사어 list 단어에 포함되는 경우, 해당 단어들만 편집거리 계산

→ 형태 변형어 원형에 가장 가까운 단어를 추출하기 위해, 두 list를 crosscheck 하여 겹치는 단어 분석

분석 단계

Idea 1) Str 길이가 같은 부분은 음소 단위로 편집거리를 계산한다.

Idea 2) Str 길이가 다른 삽입or삭제 음절의 경우, 해당 자리를 음소 단위로 편집거리를 계산하여 더한다.

Idea 3) 의미 유사어와 중복되는 음절의 개수만큼 편집거리에서 뺀다.

예시)

형태 유사어 list = {사실, 경우, 강아지는, 강아지, 카페에, 놀러, 털, 혼자, 계속, 반려견}

의미 유사어 list = {강아지옷, 이벤트, 퓨어메라, 강아지간식, 강아지, 강아지옷만들기, 애견용품}

# 진행 사항 - 특허 아이디어 발딩

+

4-1) 의미 유사어가 형태 유사어 list 단어에 포함되는 경우, 해당 단어들만 편집거리 계산

Idea 1) Str 길이가 같은 부분은 음소 단위로 편집거리를 계산한다.

⇒ 땡땡이 vs. 강아지는  $\rightarrow 2 + 2 + 1$

Idea 2) Str 길이가 다른 삽입or삭제 음절의 경우, 해당 자리를 음소 단위로 편집거리를 계산하여 더한다.

⇒ 는 “ㄴ/ㅡ/ㄴ” 음소 단위 편집거리 3 더하기  $\rightarrow 2 + 2 + 1 + 3 = 8$

Idea 3) 의미 유사어와 중복되는 음절의 개수만큼 편집거리에서 뺀다.

⇒ 강아지 음절 개수 3 빼기  $\rightarrow 8 - 3 = 5$

Aa 형태 변형어	≡ 의미 유사어 중복	≡ 형태 유사어	≡ 편집거리 계산	≡ 중복 음절 개수 빼기	≡ 최종 편집거리
땡땡이	중복 x	사실	$3 + 3 + 2 = 8$		
땡땡이	중복 x	경우	$2 + 3 + 2 = 7$		
땡땡이	강아지 중복	강아지는	Idea 1 Idea 2 $2 + 2 + 1 + 3 = 8$	Idea 3 $8 - 3$	5
땡땡이	강아지 중복	강아지	$2 + 2 + 1 = 5$	$5 - 3$	2
땡땡이	중복 x	카페에	$3 + 3 + 2 = 8$		
땡땡이	중복 x	눌러	$3 + 3 + 2 = 8$		

# 진행 사항 - 특허 아이디어 빌딩



## 5) 최종 유사어 및 편집거리 분석표 제공

→ 최종 편집거리가 가장 낮은

"강아지" 를 최종 유사어로 제공

Aa 형태 변형어	≡ 의미 유사어 중복	≡ 형태 유사어	≡ 편집거리 계산	≡ 중복 음절 개수 빼기	≡ 최종 편집거리
댕댕이	중복 x	사실	$3 + 3 + 2 = 8$		
댕댕이	중복 x	경우	$2 + 3 + 2 = 7$		
댕댕이	강아지 중복	강아지는	$2 + 2 + 1 + 3 = 8$	$8 - 3$	5
댕댕이	강아지 중복	강아지	$2 + 2 + 1 = 5$	$5 - 3$	2
댕댕이	중복 x	카페에	$3 + 3 + 2 = 8$		
댕댕이	중복 x	놀러	$3 + 3 + 2 = 8$		

## 4) 추출된 유사어들의 편집거리 계산

### 4-2) 의미 유사어가 형태 유사어 list 단어에 포함되지 않는 경우

#### 1) 형태 유사어 list에 편집거리 적용하여, 편집거리가 가장 낮은 유사어 제공

Idea 1) Str 길이가 같은 부분은 음소 단위로 편집거리를 계산한다.

Idea 2) Str 길이가 다른 삽입or삭제 음절의 경우, 해당 자리를 음소 단위로 편집거리를 계산하여 더한다.

Aa 형태 변형어	≡ 형태 유사어	≡ 편집거리 계산	≡ 최종 편집거리
댕댕이	사실	$3 + 3 + 2$	8
댕댕이	경우	$2 + 3 + 2$	7
댕댕이	강아지는	$2 + 2 + 1 + 3$	8
댕댕이	강아지	$2 + 2 + 1$	5
댕댕이	카페에	$3 + 3 + 2$	8

[('댕댕이삼계탕', '강아지삼계탕', '강아지옷', '퓨어메라', '이벤트', '강아지간식', '강아지', '강아지옷 만들기', '애견용품']

#### 2) 최종 결과에서 상위 10개 의미 유사어 함께 제공



# 차주 계획

1. 특허 아이디어 TEST – 정확도 향상을 위한 전처리 개선
2. 특허 명세서 초안 작성



감사합니다 :)  
빅데이터미네이터